Revisiting Metalevel Argumentation

Sanjay Modgil
Department of Infomatics, King's College London
sanjay.modgil@kcl.ac.uk

Abstract

This paper revisits a program of research work on metalevel argumentation conducted jointly with Trevor Bench-Capon. After a brief review of this research, I then discuss some potential uses of the metalavel approach. Specifically, I argue that one of the key benefits of the abstract argumentation paradigm is its potential for bridging between computational models of argumentation and human models of reasoning and debate, and that metalevel argumentation can play a key role in facilitating this bridging.

1 Introduction

I've known Trevor Bench-Capon since the first day I met him, and ever since have been glad to count him amongst my friends and academic collaborators. My early interest in argumentation owes much to his guidance and friendship. In 2003 John Fox offered me a postdoctoral position on the ASPIC project, and I enthusiastically accepted the opportunity of returning to the themes of my PhD – logic, reasoning and conflict – after a four year hiatus in medical informatics. It was at the project kick-off meeting, in Albi France, when I first met Trevor. My first impression, on the opening day of that meeting, was of a charmingly anarchic figure whose incisive contributions belied a look of indifference. As that first day drew to a close, I remember being somewhat underwhelmed by the prospect of the evening meal. After all, a Spanish philosopher once wrote that all human beings long for "the eternal persistence of consciousness", but my experience had been that this was not an easy state to maintain, especially when dining out with computer scientists. But conscious I remained, on that particular evening, for I discovered that many in the argumentation community constituted an altogether more entertaining breed of 'computer scientist', and chief amongst those who had me wide mouthed with laughter and slack-jawed with alcohol (his appetite then was legendary, as well as infectious), was a certain Trevor Bench-Capon. From that night on, I relished project meetings and the chances they afforded to converse with this cultured and witty colleague; and I say "colleague", since it is a testament to his lack of pretension that he made a neophyte like me feel like a colleague. Since then our friendship has grown, and as I know to be the case with many other junior researchers, I have greatly benefitted from his advice, guidance and support (especially when at one point I felt that the chances of a permanent academic position were hopeless).

I have also enjoyed many fruitful academic collaborations with Trevor. In 2007, I asked him to comment on an early draft of what was to become my paper on Extended Argumentation Frameworks (*EAF*s) [15]. His comments were very helpful, and in particular, he suggested an idea that subsequently evolved into our work on metalevel argumentation [7], [16], [17], [18]. Essentially, the idea of metalevel argumentation is that given an object-level argumentation framework (such as a Dung framework (AF) [11] or an *EAF*), one can consider metalevel arguments that can be explicitly categorised according to the types of claim made about the arguments and their relations in the object level framework. These metalevel arguments can then themselves be related by an attack relation in a Dung framework, where this metalevel attack relation satisfies constraints imposed by the claim based categorisation. One can then show a correspondence between the object level framework and its metalevel formulation, such that the justified arguments of the object level framework can be computed directly from its metalevel formulation[1]. In this way, the full range of theoretical and practical results and techniques for AFs can now be inherited by their various extensions and developments, including *EAF*s, frameworks with support relations [2], recursive attacks [4], collective attacks [21], etc.

In this paper I will explore uses and applications of meta-argumentation that go beyond those described in our above mentioned papers on metalevel argumentation. Section 2 briefly reviews Dung's argumentation theory and metalevel argumentation as formalised in [18]. Section 3 then proposes the use of metalevel argumentation as a formalism for bridging between computational and human models of argument, whereby networks capturing interactions between human authored statements and arguments can be mapped to metalevel argumentation frameworks, so that evaluation of the latter under Dung's standard semantics can provide dialectical guidance to human users, and more sophisticated feedback prompting human users to submit dialogical moves that render explicit, information that is implicitly encoded in the relations holding between statements and arguments. The paper then concludes in Section 4.

2 Background: Abstract and Metalevel Argumentation Frameworks

2.1 Abstract Argumentation Frameworks

Many applications of argumentation build on Dung's seminal theory [11] and its various developments. A Dung *argumentation framework* (*AF*) consists of a binary conflict based *attack* relation R on a set A of arguments. Then, $x \in A$ is said to be *acceptable* w.r.t. $S \subseteq A$ iff $\forall y \in A$ such that yRx, implies $\exists z \in S$ such that zRy. This basic principle, whereby x is defended (or reinstated) against an attack by y, if some z attacks y, underpins evaluation of the winning/justified arguments of an *AF* in the following way:

[1]The idea of metalevel argumentation first appeared in [16], and was subsequently adopted by Boella et.al [9]. In their work, formal correspondences with object level frameworks are not shown.

Definition 1 Given an *AF* (A, R), Let $S \subseteq A$ be conflict free iff $\forall x, y \in S, (x, y) \notin R$, and let $S \subseteq A$ be an admissible extension iff S is conflict free and all arguments in S are acceptable w.r.t. S. The status of arguments is then evaluated w.r.t. extensions defined under different semantics:

Let S be an admissible extension of (A, R).

• S is *complete* iff S contains all arguments in A which are acceptable w.r.t S; *grounded* iff S is the minimal (w.r.t. set inclusion) *complete* extension; *preferred* iff S is a maximal *complete* extension, and *stable* iff $\forall y \notin S, \exists x \in S$ s.t. $(x, y) \in R$

• For $s \in \{$complete, preferred, grounded, stable$\}$:
If $x \in A$ is in at least one, respectively all, s extension(s) of (A, R), then x is said to be credulously, respectively sceptically, justified under the s semantics.

Dung's theory has been developed in a number of directions. Some works formalise collective attacks between *sets* of arguments [21]. In other works, the success of an attack from x to y, as a *defeat* by x on y, is contingent on y not being preferred to x according to some given preference relation on $\mathcal{A}$ [1], or the value promoted by y not being ranked higher than the value promoted by x, according to a given ordering on values [5]. [15]'s *Extended Argumentation Framework* (EAF) then extended Dung's framework to include arguments that attack attacks. EAFs thus accommodate argumentation based reasoning *about* possibly conflicting preference information, values, and value orderings, within the argumentation framework itself. [4] generalise the idea of attacks on attacks to recursive attacks on attacks, while a number of works also augment Dung's framework to include a *support* relation on arguments [2],[22], and weights on attacks [12].

2.2 Metalevel Argumentation Frameworks

Metalevel Argumentation Frameworks (*MAF*s) [18] categorise meta-arguments according to the claims they make *about* object level arguments and their properties and relations. These meta-arguments are organised into a Dung AF whose meta-attack relation obeys constraints imposed by the claim based characterisation.

Definition 2 A MAF is a tuple $\Delta_{\mathcal{M}} = (\mathcal{A}, \mathcal{R}, \mathcal{C}, \mathcal{L}, \mathcal{D})$, where $(\mathcal{A}, \mathcal{R})$ is a Dung AF, and:

- $\mathcal{L}$ is a language that includes a countable set of constant symbols and predicates. The set *wff*$(\mathcal{L})$ is defined by the following BNF (x, x_i range over constant symbols or variables)[2]:
  $\mathcal{L}: X ::= x, \{x_1, \ldots, x_n\} \mid justified(X) \mid rejected(X) \mid attack(X, X') \mid defeat(X, X') \mid preferred(X, X') \mid support(X, X') \mid unsupported(X, X')$
- The claim function $\mathcal{C}$ is defined as $\mathcal{C}: \mathcal{A} \mapsto 2^{wff(\mathcal{L})}$
- $\mathcal{D}$ is a set of constraints on $\mathcal{R}$ of the form:
  if $l \in \mathcal{C}(\alpha)$ and $l' \in \mathcal{C}(\beta)$ then $(\alpha, \beta) \in \mathcal{R}$

[2] In [18] $\mathcal{L}$ also includes val, val_pref, $audience$ and *wff* constructed from these predicates.

• $\mathcal{R}$ is said to be *defined by* $\mathcal{D}$ if whenever $(\alpha, \beta) \in \mathcal{R}$ then the claims of α and β satisfy the antecedent of some constraint in $\mathcal{D}$.

• The extensions and justified arguments of $\Delta_{\mathcal{M}}$ are the extensions and justified arguments of ($\mathcal{A}$, $\mathcal{R}$).

Henceforth, we may use abbreviations $j, r, d, p \ldots$ for $justified$, $rejected$, $defeat$, $preferred$ etc., and may also denote an argument by the claim it makes. For example, $j(x)$ may denote the meta-argument claiming x is justified.

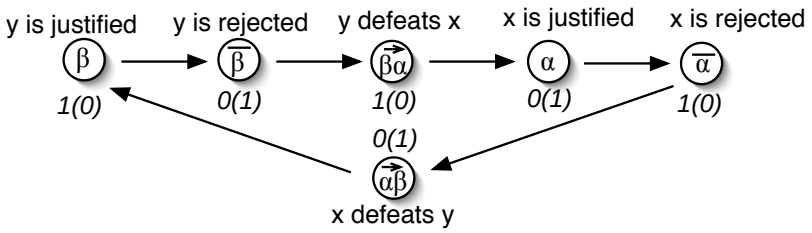

Figure 1: The MAF characterisation of a Dung $AF\ x \rightleftarrows y$

Consider now a given object level AF, (A, R). Then the existence of an argument $x \in A$, gives rise to a meta-argument $\alpha \in \mathcal{A}$ of the form 'there is an $x \in A$ that is an admissible extension of (A, R)', supporting the claim that 'x *is justified*'. The existence of an object level attack yRx, constitutes a meta-argument $\overrightarrow{\beta\alpha}$ = 'y attacks x' supporting the claim 'y *defeats* x'. Since the justified status of x in the object level framework is challenged by a defeat on x, then $\overrightarrow{\beta\alpha}$ attacks α at the metalevel, and so we have the following constraint on the meta-level attack relation $\mathcal{R}$ (V, W, X, Y, Z will henceforth range over *wff* of $\mathcal{L}$):

D1 : if $d(Y, X) \in \mathcal{C}(\gamma)$ and $j(X) \in \mathcal{C}(\alpha)$ then $(\gamma, \alpha) \in \mathcal{R}$

y does not defeat x if y is rejected, and so $\overrightarrow{\beta\alpha}$ is attacked by a meta-argument $\overline{\beta}$ claiming 'y *is rejected*'. However, y does defeat x if y is justified, and so β claiming 'y *is justified*' attacks $\overline{\beta}$. We thus have the following metalevel constraints:

D2 : if $d(Y, X) \in \mathcal{C}(\gamma)$ and $r(Y) \in \mathcal{C}(\beta)$ then $(\beta, \gamma) \in \mathcal{R}$

D3 : if $j(X) \in \mathcal{C}(\alpha)$ and $r(X) \in \mathcal{C}(\beta)$ then $(\alpha, \beta) \in \mathcal{R}$

Fig 1 shows the MAF characterisation of a Dung $AF\ x \rightleftarrows y$ (together with the two labellings – the second in brackets – identifying the two preferred extensions). In [18] the following correspondence is shown:

> Let $\Delta = (A, R)$, $\Delta_{\mathcal{M}}$ its MAF ($\mathcal{A}, \mathcal{R}, \mathcal{C}, \mathcal{L}, \mathcal{D}$), where $x \in A$ iff $j(x), r(y) \in \mathcal{A}$, $(y, x) \in R$ iff $d(y, x) \in \mathcal{A}$, and $\mathcal{R}$ is defined by $\{D1, D2, D3\}$. Then x is a justified argument of Δ iff $j(x)$ is a justified argument of $\Delta_{\mathcal{M}}$ (under any semantics).

Note that in the spirit of Dung's AFs, MAFs adopt an abstract level approach in that they leave open the question of how meta-arguments might be formally constructed (in some meta-logic), and specify at the abstract level: 1) a function that maps meta-arguments to their claims, expressed in a language (that may be) distinct from that in which the arguments are formally constructed; 2) constraints on the attack relation defined in terms of these claims.

In [18], preference based AFs (PAFs) [1], value based AFs [5], hierarchical *EAF*s [15], and frameworks with collective attacks [21] are all characterised as metalevel Dung frameworks, and similar correspondences are shown. In addition, [17] shows how argument accrual can also by formalised in MAFs, so integrating accrual and dialectical modes of argumentation. An example of the correspondence shown in [18], involves characterising attacks on attacks in *hierarchical EAF*s, as metalevel attacks on arguments claiming object level defeats. Recall that in [15], it may be that in an *EAF*, x attacks y, but arguments may also attack attacks, so that z may attack the attack from x to y, indicating that z is an argument claiming that y is preferred to z. Hence, if z is a justified argument, then the attack from x to y fails to succeed as a defeat. Fig 2 shows the metalevel characterisation of the object level attack on an attack.

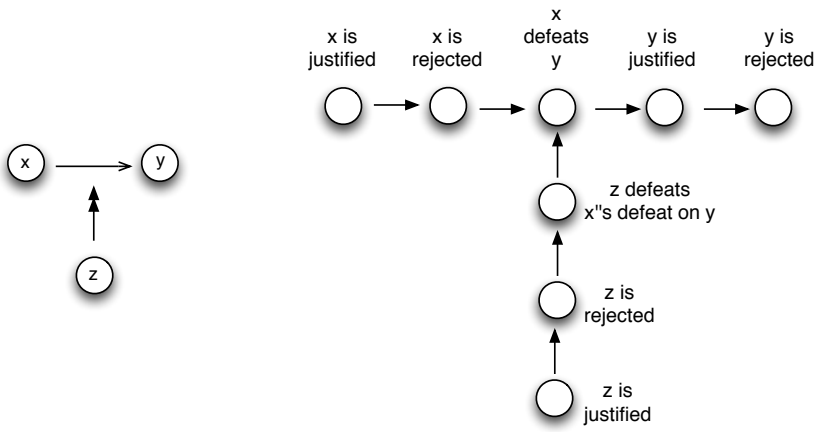

Figure 2: The EAF attack on an attack is shown on the left. The metalevel characterisation is shown on the right, with the claims of the metalevel arguments also shown.

3 Metalevel Argumentation: Bridging between Human and Formal Models of Argumentation

In our papers on metalevel argumentation we discuss various uses of metalevel argumentation frameworks. An obvious advantage of the approach is that MAF characterisations of object level formalisms allows for application of the full range of results

and techniques developed for Dung *AF*s to be applied to the various object level developments of *AF*s (e.g.labelling algorithms and argument game proof theories [19] for Dung frameworks can now be applied to extensions of Dung's framework). Furthermore, MAFs provide a unifying formalism for integrating and further extending the various developments of Dung argumentation (e.g., integrating preference and value-based argumentation as shown in [18]). Specific uses of metalevel argumentation have been described in: [16], in which dialogue games for metalevel formulations of value-based argumentation were shown to have advantages over games defined for object level *VAF*s [6]; [7] in which argumentation about cases and their uses as precedents in legal reasoning can be modelled by dialogues defined over MAFs, and ; [3], in which MAFs are used for reasoning about firewall policies.

In what follows, I will discuss a further use of metalevel argumentation; in particular, as a bridging formalism between human and computational models of argument. To provide an appropriate context for the ensuing discussion it is worth reviewing what can be considered to be one of the key reasons for why argumentation has emerged as a prominent logic-based paradigm for reasoning under uncertainty and conflict [3].

3.1 Argumentation: The Added Value

The continuing impact of Dung's theory can be attributed to its level of abstraction and characterisation of non-monotonic inference relations in terms of general and intuitive principles. One is free to choose a logic and define what constitutes an argument and attack for that logic[4]. Then, given the arguments and attacks defined (instantiated) by a possibly inconsistent set of *wff* in that logic, one evaluates the justified arguments. The claims of these arguments then identify the non-monotonic inferences from the set of instantiating *wff*. Thus, abstract argumentation defines non-monotonic inference relations for instantiating monotonic logics. Furthermore, existing non-monotonic logics (e.g. logic programming, default, auto-epistemic and defeasible logic) can be given argumentation-based characterisations ([11],[13]), in the sense that the inferences defined though instantiation and evaluation of justified arguments correspond to the inference relations of the instantiating non-monotonic logic.

The fact that reasoning in existing non-monotonic logics can thus be characterised, testifies to the generality of the principle whereby one argument defends another from attack; a principle that is also both intuitive and familiar in human modes of reasoning, debate and dialogue. Indeed, recent ground breaking work in cognitive science argues that the human capacity for reasoning evolved primarily in order to assess and counter the claims and arguments of interlocutors in social settings [14]. Argumentation theory thus provides a *language independent* characterisation of both human and logic-based reasoning in the presence of uncertainty and conflict, through the abstract dialectical modelling of the process whereby arguments can be moved to attack and defend other arguments. The theory's value [5] can therefore in large part be attributed

[3] Five of the top ten most cited articles in the Journal of Artificial Intelligence, between 2007 and 2012, were on the topic of argumentation.

[4] Although as shown in [20], any given choice may be ill-conceived in that the resulting instantiated argumentation framework does not satisfy desirable properties (rationality postulates).

[5] Note that some of the material in this section draws from material presented in Section 21.1.2 of [24]

to its explanatory potential for making non-monotonic reasoning processes inspectable and readily understandable for human users, and its underpinning of dialogical and more general communicative interactions involving reasoning in the presence of uncertainty and conflict, where such interactions may be between heterogeneous agents (i.e., computational and/or human). Thus, computational reasoning processes can be informed by argumentation-based characterisations of human reasoning and interaction, and the reasoning processes of humans can be informed by argumentation-based characterisations of computational reasoning.

However, in order that argumentation can provide such a bridging role between computational and human reasoning, one requires development of models that account for human reasoning and argument as conducted in practice. Note that in this view, the oft heard critique aimed at abstract argumentation formalisms needs to be reformulated. This critique argues that in order to justify the addition of concepts and constructs extending AFs at the abstract level, one should relate these extensions to the underlying logical features that they abstract from; otherwise the abstraction counts for nothing, since it is an abstraction of nothing. However, the aforementioned requirement suggests that these various extensions can also be justified according to whether they accommodate modes of human reasoning and argument. Hence the critique might be better re-cast as a challenge for any such developed extension: *either* demonstrate that the extension provides a framework for instantiation by some underlying logical formalism, *or* demonstrate that the abstraction developed intuitively accommodates modes of human argumentation.

Given abstract extensions of AFs that do accommodate human modes of reasoning and argument, then in order to facilitate the bridging role of argumentation requires formally relating these extensions to the computational logic-based models of argument exemplified by Dung frameworks. This is where I believe metalevel argumentation can play an important role.

3.2 Dialectical Evaluation of Arguments Authored by Humans

To illustrate, a key anticipated use of computational argumentation is in informing human debate and argument so that interlocutors are guided by logical rational principles. The idea is that human dialogue and debate on online and offline tools (e.g., *Rationale* [8] and the tools developed at Dundee University that are reviewed in [24]) are mapped to Dung frameworks, and evaluated under Dung's various semantics. The provision of this evaluative functionality would: 1) ensure that the assessment of arguments is formally and rationally grounded; 2) enable humans to track the status of arguments so that they can be guided in which arguments to respond to; 3) enable 'mixed' argumentation integrating both computational and human authored arguments.

To achieve this functionality requires that tools accommodate interactive reasoning and debate as conducted in real life, where interlocutors sometime exchange complete arguments, incomplete arguments (enthymemes), individual statements (that may combine to form arguments), or indeed questions and challenges that prompt responses. They may also debate the relative strength of arguments, or whether one argument does indeed constitute a valid attack on another argument, or submit arguments supporting other arguments or accruing for a given claim, etc. The contributions of interlocutors

then need to be organised into Dung frameworks and evaluated.

Let me illustrate with a simple example of a fictional exchange between myself and Trevor, an entirely appropriate advocate given his quickness to provoke and cajole one into a 'robust' argument, which is not to imply reproach; after all, mediocrity thrives on a diet of consent, and mediocrity is an anathema to Trevor as good football is an anathema to Portsmouth FC:

Sanjay : "The information about Tony Blair's affair should not have been published ($\neg pub$) because it was private information pr." (A)

Trevor : "So what ? Just because it is private, why should that mean it should not be published ?"

Sanjay : "Well, he's also no longer a public figure ($\neg p\_f$)." (B)

Trevor : "So what?"

Sanjay : "Well, also the information is not in the public interest ($\neg p\_int$)". (C)

Trevor : "But Blair has been appointed as UN envoy for the Middle East (un_env), and so the information is in the public interest". (D)

The exchange can be characterised in terms of an argument A, supported by arguments B and C, where C is then attacked by D. Of course, in this example the support relation essentially expresses a sub-argument relation, in that the support of B and C for A can be modelled as a single argument whose claim is $\neg pub$, inferred from the defeasible rule $pr \wedge \neg p\_f \wedge \neg p\_int \Rightarrow \neg pub$ and premises $pr, \neg p\_f$ and $\neg p\_int$. However, the point is to capture the arguments, statements, and their relations as interlocutors might present them.

While I acknowledge that the notion of support has been ascribed a variety of interpretations in the literature [6], it is often the case that illustrative examples suggest the sub-argument interpretation is the one implied (as in the above case), and this is confirmed by the fact that an attack on an argument X is extended to an attack on arguments supported by X [2]. Hence, the metalevel characterisation of this notion of support is that if X supports Y, and X is rejected, then Y is unsupported by X, and so rejected. Consider now the abstract framework shown in Figure 3 in which the arguments in the above dialogue are related by attack and support relations, and the associated metalevel interpretation. Then, one can apply standard techniques (such as labelling algorithms [19]) to evaluate the winning arguments; in this case providing me with feedback that my argument A is losing.

3.3 Prompting Dialogical Moves

In what follows I speculate on more sophisticated uses of metalevel argumentation in bridging between human and computational models of argument. Specifically, the

[6] Another reasonable interpretation is that X supports Y if X's claim α is a premise or conclusion of a rule in Y so that X is an additional argument for α, i.e., we have an instance of arguments 'accruing' for α: X and the sub-argument Y' of Y that concludes α.

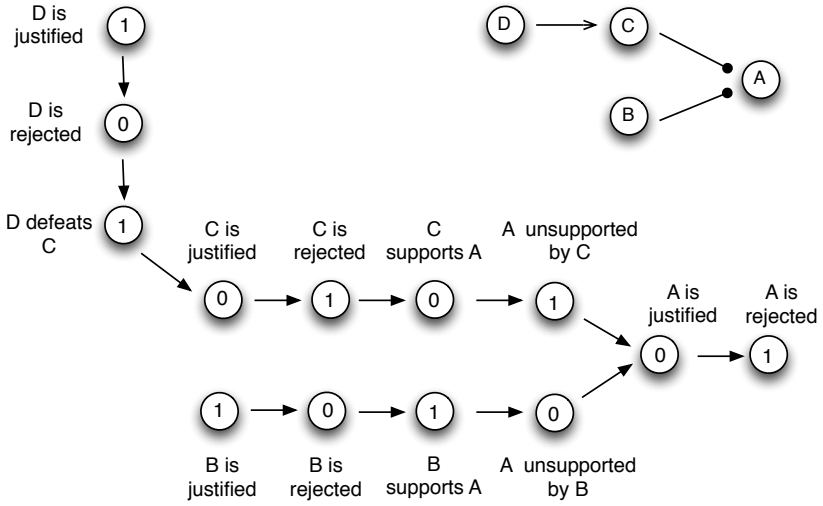

Figure 3: The framework above contains support relations (links with swollen ends). Its metalevel characterisation is shown below with the associated admissible/grounded labelling; 0 for losing (rejected) and 1 for winning (justified).

previous section's use of metalevel frameworks, in providing dialectical feedback, did not rely on the internal structure of the metalevel arguments. Rather, the feedback relied on their organisation into a Dung graph, which in turn relied only on the claims these meta-arguments make about object level frameworks. As discussed in Section 2.2, we did not in [18] formally define instantiation of meta-arguments, but appealed to informal notions such as α is a meta-argument of the form 'there is an $x \in A$ that is an admissible extension of (A, R)', supporting the claim that x *is justified*, where this claim is associated with the argument at the abstract level (and so does not commit to the specifics of the language used for instantiating the meta-argument). I will now discuss how the following variation of the previous section's dialogue illustrates how future work on more formal instantiations of meta-arguments (e.g., of the type described in [25]) will enable more sophisticated user feedback.

> Suppose Sanjay argues that "Tony Blair is no longer a public figure, the information about his affair is not in the public interest, and the information is private, so the information should not be published" (X). Trevor counter-argues with "but Blair is UN envoy for the Middle East" (Y). Sanjay then counters by asking "why is this an argument against prohibiting publication ?" (Z). Trevor strokes his finely coiffeured goatee, and responds with "because his appointment as UN Middle East envoy implies that the information about his affair is in the public interest" (V).

This example raises a number of issues:

The distinct roles of attack in argumentation. Firstly note that I counter Trevor's attack on X, by Y, by questioning the validity of the attack. This question can then be represented as an attack on Y's attack on X, but *not* an EAF meta-attack [15], where an attack from Z to Y's attack on X invalidates the *dialectical* use of Y as a counter-argument to X, given that Z is an argument claiming that X is preferred to Y. Rather, I question the *declarative* basis for the attack.

To elaborate, object level attacks play two roles. Firstly, that Y attacks X is an abstract, declarative representation of the mutual incompatibility of the claim of the attacking argument and some element in the attacked argument. Secondly, the attack abstractly characterises the dialectical, procedural use of Y as a counter-argument to X. Definition 1's notion of a conflict free set accounts for the former declarative denotation, whereas the notion of acceptability of arguments accounts for the dialectical use of attacks. The question that then naturally arises is how can one question, in a formal *logical* context, the declarative basis of an attack from Y to X, since to do so would be to question the fundamental logical principle that a formula (i.e., the claim of Y) and its negation (i.e., a premise or conclusion of a rule in X) are in conflict ? Clearly one cannot do so in a logical context[7], but in informal human contexts in which enthymemes – arguments in which information is omitted – are commonplace, one can question the declarative rationale for an attack in those cases where the missing information is that which would provide such a rationale.

Attacking the declarative basis of attacks in enthymemes. In our example, Y is just such an enthymeme. The very fact that Y is moved as an attack on X (as indicated by the qualifying "but"), but the attack is not explicitly targeted, is indicative of an incomplete rule of the form 'if someone is a UN envoy for the Middle East then (s)he is ...', where the missing information is some proposition that negates an element in X. Note that if explicitly targeted, as for example in Trevor's argument D in the previous section's dialogue, it is clear that "the information is in the public interest" negates a premise in Y. In such a case it makes no sense to question the declarative basis of this attack; rather the rule $un\_env \Rightarrow p\_int$ in D would be targeted by an attack on D.

Metalevel prompting of dialogical moves. Suppose the locutions in this section's dialogue are represented together with the relations between them. Figure 4 shows the incremental construction of the framework[8], and the associated MAF.

1. My assertion of X, is associated with assertion of meta-arguments α claiming $justified(x)$ and $\overline{\alpha}$ claiming $rejected(x)$.
2. Now, as suggested earlier, let us suppose construction of metalevel arguments using a metalogic of the type described in [25]. Trevor's assertion of the attacking Y yields meta-arguments β and $\overline{\beta}$ respectively claiming $justified(y)$ and $rejected(y)$. It also yields the premise $attack(y, x)$, which together with the defeasible rule '$attack(y, x) \Rightarrow defeat(y, x)$' yields the meta-argument $\overrightarrow{\beta\alpha}$

[7] Although this is not strictly true of approaches that make use of contrary relations (e.g., ABA [10] and ASPIC+ [20]) which generalise negation. But then in these approaches such relations cannot be subject to argument.

[8] I hesitate to call this an 'argument' framework; rather it is a dialogue framework given that the locutions represented by the nodes do not strictly equate with arguments construed as reasons in support of a claim; on the other hand the relations between the nodes are those we are familiar with in argument frameworks

claiming $defeat(y, x)$. In other words, the launching of the attack allows one to defeasibly conclude that it succeeds, dialectically, as a defeat.

3. Now, a challenge on the premise of $\overrightarrow{\beta\alpha}$ is suggested, i.e., 'why $attack(y, x)$?', shifting the burden of proof on Trevor to justify the attack. This is exactly the kind of dialogical move one sees in persuasion dialogues [23], where a *why* locution challenging a premise is effectively interpreted as an attack on the premise, in the sense that the burden of proof is then on the interlocutor submitting the premise to provide an argument for that premise. The metalevel arguments associated with this challenge are γ = 'there exists a challenge z' supporting the claim $justified(z)$ which attacks $\overline{\gamma}$ claiming $rejected(z)$, which in turn attacks $\overrightarrow{\gamma_{\overrightarrow{\beta\alpha}}}$ claiming $defeat(z, attack(y, x))$, which in turn attacks $\overrightarrow{\beta\alpha}$ on its premise $attack(y, x)$. Notice the contrast with the meta-argument in Figure 2, representing the EAF attack on the dialectical success of the attack from x to y; it's claim thus being $defeat(z, defeat(x, y))$[9].
4. Finally, Trevor responds to the challenge by supplying the missing information, i.e., the rule V = $un\_env \Rightarrow p\_int$ that resolves the issue of which of X's premises is targeted by Y.

What the above suggests, is that the metalevel characterisation of the dialogue as a MAF, prompts submission of dialogical moves (Z) that then serve to render explicit, implicit information in enthymemes (V), so that this information is made available for debate. For example, I might then respond to Trevor by undercutting the rule $un\_env \Rightarrow p\_int$ (and thus the argument composed from Y and V). To reiterate, contrast this section's dialogue with that in the previous section, in which the targeted attack by D would yield a meta-argument composed from premise stating that 'D's claim negates the claim of C', which then strictly (rather than defeasibly) implies that 'D attacks C', which in turn defeasibly implies that 'D defeats C'.

From abstract argumentation to computational knowledge: Clearly, much of what is described above awaits a more formal analysis. However, I hope the intuitions are convincing enough to suggests a number of interesting research issues. Indeed, the above examples illustrate a more general research goal that I am currently pursuing; the exploitation of abstract relations between locutions/arguments, to, as it were, induce further information for updating computational knowledge. A simple example of this is suggested by the previous section's dialogue: the moving of B and C in response to Trevor's 'so whats' is indicative of B and C's (sub-argument) support for A, and so can be seen as inducing the rule $pr \wedge \neg p\_f \wedge \neg p\_int \Rightarrow \neg pub$. Also, in this section's dialogue, we showed how further dialogical moves can be prompted, serving both the overall dialectical goal of the dialogue, and in so doing eliciting the rule $un\_env \Rightarrow p\_int$.

[9] Which, now that we take into account the logical structure of the meta-argument $\overrightarrow{\beta\alpha}$, could then be seen as either rebutting the claim $defeat(y, x)$ or undercutting of the rule $attack(y, x) \Rightarrow defeat(y, x)$

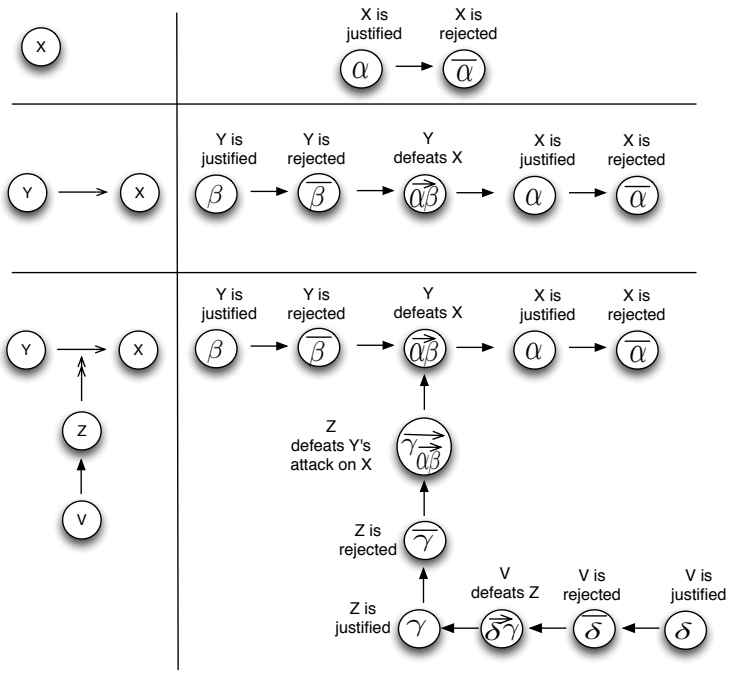

Figure 4: The frameworks on the left are those incremented during the dialogue, while the associated increments to the MAF are shown on the right.

4 Conclusions

In this paper I have briefly reviewed a program of research work on metalevel argumentation conducted jointly with Trevor Bench-Capon. I have also looked forward to what I believe are promising research directions that are complementary to commonly accepted views of the practical benefits that the argumentation paradigm has to offer. Essentially, I've suggested that metalevel argumentation can bridge between human modes of reasoning and debate, and computational models of argumentation. Human submitted arguments and statements can be captured in frameworks that can be represented as metalevel Dung frameworks. Further debate can then be informed by dialectical evaluation of these frameworks, and the prompting of further dialogical moves that seek to clarify and make available implicit information that can then be further interacted with.

Finally, I would like to emphasise the pleasure it gives me in being able to contribute to this Festschrift, especially in view of what I witnessed Trevor endure in these last few years. This paper is a testament not only to the importance of the metalevel paradigm that we developed, but also to my delight in seeing him triumph, against the odds.

References

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[2] L. Amgoud, C. Cayrol, M. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23(10):1062–1093, 2008.

[3] A. Applebaum, A.R. Syed, K. N. Levitt, S. Parsons, J. Rowe, and E. Skar. Firewall configuration: An application of multiagent metalevel argumentation. In *Proc. 9th International Workshop on Argumentation in Multiagent Systems (ArgMAS)*, 2012.

[4] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. AFRA: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning*, 52(1):19 – 37, 2011.

[5] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[6] T. J. M. Bench-Capon, S. Doutre, and P. E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171(1):42–71, 2007.

[7] T. J. M. Bench-Capon and S. Modgil. Case law in extended argumentation frameworks. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ICAIL '09, pages 118–127, 2009.

[8] T. Berg, T. van Gelder, F. Patterson, and S. Teppema. *Critical Thinking: Reasoning and Communicating with Rationale*. Amsterdam: Pearson Education Benelux, 2009.

[9] G. Boella, D.M. Gabbay, L. Van de Torre, and S. Villata. Meta-argumentation modelling 1: Methodology and techniques. *Studia Logica*, 93:297–355, 2009.

[10] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[11] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[12] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457 – 486, 2011.

[13] G. Governatori and M. J. Maher. An argumentation-theoretic characterization of defeasible logic. In *Proc. 14th European Conference on Artificial Intelligence*, pages 469–473, 2000.

[14] H. Mercier and D. Sperber. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–747, 2011.

[15] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

[16] S. Modgil and T. J. M. Bench-Capon. Integrating object and meta-level value based argumentation. In *Computational Models of Argument: Proceedings of COMMA 2008*, pages 240–251, 2008.

[17] S. Modgil and T. J. M. Bench-Capon. Integrating dialectical and accrual modes of argumentation. In *Computational Models of Argument: Proceedings of COMMA 2010*, pages 335–346, 2010.

[18] S. Modgil and T. J. M. Bench-Capon. Metalevel argumentation. *Journal of Logic and Computation*, 21(6):959–1003, 2011.

[19] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G. Simari, editors, *Argumentation in AI*, pages 105–129. Springer-Verlag, 2009.

[20] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195(0):361 – 397, 2013.

[21] S. H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Proc. 3rd Int. Workshop on Argumentation in Multi-agent Systems*, pages 54–73, 2006.

[22] N. Oren and T. J. Norman. Semantics for evidence-based argumentation. In *Computational Models of Argument: Proceedings of COMMA 2008*, pages 276–284, 2008.

[23] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.

[24] S. Modgil and F. Toni et.al. Chapter 21: The added value of argumentation. In Sascha Ossowski, editor, *Agreement Technologies*, pages 357–403. Springer Netherlands, 2013.

[25] M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In *Proc. Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 560–567, 2005.